

REMARKS

Claims 1 and 3-29 were pending in the application. Claims 1 and 3-29 stand rejected. Claims 1 and 3-29 remain in the application.

Claims 1, 3-8, 10, and 23-29 stand rejected under 35 U.S.C. 103(a) as being unpatentable over Qian et al., US Patent 6721454 (hereafter referred to as "Qian 454") and further in view of Ratakonda, US Patent 5956026. The rejection stated:

"As in Claims 1, 24, and 27-29, Qian et al. teaches a method and computer storage medium with instructions for obtaining unstructured video frames ("A video sequence 2 is input", Column 2, lines 64-65), generating segments from the shot boundaries based on the color dissimilarity between consecutive frames ("A color histogram technique may be used to detect the boundaries of the shots", Column 3, lines 42-43), extracting a set by processing pairs of segments ("the global motion of the video content is estimated 8 for each pair of frames in a shot", Column 3, lines 59-61) for their visual dissimilarity and temporal relationship, generating a feature including metrics of temporal separation between segments of the respective pair and accumulated duration of segments of the pair (temporal and spatial phenomena), and merging the video segments by applying a probabilistic analysis to the extracted set to represent the video structure ("each shot is summarized 16 ...events 22 are inferred from the shot summaries by a domain specific event inference model". Column 3, lines 6-8). While Qian et al. teaches extracting semantic events from unstructured video frames, they fail to show the generation of inter-segment color dissimilarity feature and inter-segment temporal relationship feature of each pair of segments as recited in the claims. In the same field of the invention, Ratakonda teaches a video event detection and segmentation merging method similar to that of Qian et al. In addition, Ratakonda further teaches the generation of inter-segment color dissimilarity feature and inter-segment temporal relationship feature of each pair of segments (Figures 1, 5 and corresponding text). It would have been obvious to one of ordinary skill in the art, having the teachings of Qian et al. and Ratakonda before him at the time the invention was made, to modify the segment generation and merging techniques taught by Qian et al. to include the processing of each pair of segments of Ratakonda, in order to obtain not only frames, but also inter-segment similarity processing. One would have been motivated to make such a combination because layered hierarchical structure would have been obtained, as taught by Ratakonda."

The rejection hinges upon a position directly contradicted by the cited references.

Claim 1 requires "extracting a feature set by processing pairs of said segments". The rejection relies on the position that Qian 454 teaches extracting a feature set by processing pairs of segments, because Qian 454 teaches the global motion of the video content is estimated for each pair of frames in a

shot. (The terms: "segment" and "shot" are both widely used in the art and have the same meaning.) The Office Action states this position in the rejection of Claim 1 and in the Response to Arguments:

"extracting a set by processing pairs of segments ("the global motion of the video content is estimated 8 for each pair of frames in a shot", Column 3, lines 59-61)' (page 3, emphasis added)

"In response to the arguments stating that Qian fails to teach descriptor of different shots are compared but not in pairs, the examiner disagrees. Qian teaches each pair of frames compared (Col. 3, lines 59 et seq.)" (emphasis added)

To take this position, the rejection, contrary to the Qian 454 and Ratakonda, incorrectly equates the terms "shot" and "segment", which each refer to a sequence of images, with the term "frame", which refers to a single image. The meanings of the terms "frame" and "shot" are defined in Qian 454:

"A video sequence includes one or more scenes which, in turn, include one or more video shots. A shot comprises a plurality of individual frames of relatively homogeneous content." (Qian 454, col. 3, lines 37-40; emphasis added; Ratakonda is in accord. See Ratakonda, col. 2, lines 13-34)

A shot or segment is not a frame. In Qian 454, shots are summarized with descriptors, such as "animal" and "tree", and the descriptors of different shots are compared, but not in pairs. (Qian 454, col. 10, line 61 to col. 12, line 9) The cited comparison of a pair of frames (Qian 454, col. 3, lines 59 et seq.) is not a comparison between shots. Qian 454 states at col. 3, lines 59-61:

"At the first level 4 of the technique, the global motion of the video content is estimated 8 for each pair of frames in a shot." (emphasis added)

The rejection, thus, is not supported by the cited references and must be withdrawn. The remaining claims are also allowable on the same grounds. The cited references do not teach or suggest extracting a feature set of features of each pair of segments.

Claim 1 requires "extracting a feature set by processing pairs of said segments, said extracting generating an inter-segment color dissimilarity feature and an inter-segment temporal relationship feature of each said pair of

segments". The rejection proposes that Qian 454 teaches "extracting a set by processing pairs of segments ("the global motion of the video content is estimated 8 for each pair of frames in a shot", Column 3, lines 59-61)". As discussed above, a shot or segment is not a frame. Thus, the determinations of motion estimates of each pair of frames in Qian 454, col. 3, lines 59-61 do not teach or suggest extracting a feature set by processing pairs of segments. The rejection also indicates:

"In addition, Ratakonda further teaches the generation of inter-segment color dissimilarity feature and inter-segment temporal relationship feature of each pair of segments (Figures 1, 5 and corresponding text)."

Figure 1 of Ratakonda does not show pairs of segments and is not discussed in terms of "each pair of segments". Figure 5 shows a finest level 76, which includes five keyframes. The rejection is apparently based upon an assumption that each of the keyframes in Figure 5 of Ratakonda corresponds to a shot (segment). This assumption has appeared to have some support. Ratakonda does state:

"Tagging [i.e. clicking on] frames in the finest level 76 results in playback of the video; for instance if the j-th keyframe is tagged at the finest level, frames between the j th and (j+1) st keyframes are played back."

(Ratakonda, col. 5, lines 51-54; see also col. 5, lines 56-59);

however, it is noted that there is no teaching or suggestion that "frames between the j th and (j+1) st keyframes" are a shot or segment. On the other hand, Ratakonda teaches to the contrary:

'A major limitation of the above schemes is that they treat all shots equally. In most situations it might not be sufficient to represent the entire shot by just one frame. This leads to the idea of allocating a few keyframes per each shot depending upon the amount of "interesting action" in the shot.' (Ratakonda, col. 1, line 64 to col. 2, line 1)

'Given total number of keyframes (user specified) 40, each shot is assigned a number of keyframes 42 depending upon the "action" within the shot, according to well known techniques.' (Ratakonda, col. 4, lines 54-57; emphasis added; also see Ratakonda Figure 2)

Ratakonda also states:

"The number of keyframes allocated to a particular shot 's', block 42, is proportionate to the relative amount of cumulative action measure within that shot." (Ratakonda, col. 6, lines 42-44; see also pruning of keyframes of some shots, Ratakonda, col. 8, lines 31-34)

Figure 5 is in accord with these quotes. Figure 5 shows a coarse level 74 of keyframes, each of which is associated with a group of finest level keyframes. One of the groups has three finest level keyframes. The other has two, that is, a pair. Figure 5 thus shows two groups of different sizes, not pairs. (See also Ratakonda, col. 5, lines 49-51.) Ratakonda does discuss use of "pairwise" clustering algorithms, but this clustering is of finest level keyframes not segments (shots). (Ratakonda, col. 9, lines 40-64) There is, thus, no teaching or suggestion of extracting a feature set of features of each pair of segments.

The cited references do not teach or suggest metrics of temporal separation between segments and accumulated duration of the segments.

Claim 1 requires "said inter-segment temporal relationship feature including metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair". The rejection proposes that Qian 454 teaches extracting a feature set by processing pairs of segments:

'for their visual dissimilarity and temporal relationship, generating a feature including metrics of temporal separation between segments of the respective pair and accumulated duration of segments of the pair (temporal and spatial phenomena), and merging the video segments by applying a probabilistic analysis to the extracted set to represent the video structure ("each shot is summarized 16 ...events 22 are inferred from the shot summaries by a domain specific event inference model". Column 3, lines 6-8).'

Qian 454 teaches shot summaries in the form of text descriptors. Some of the descriptors are labelled "'temporal descriptors". Qian 454 states:

'Temporal descriptors represent motion information related to objects and the temporal relations between them. These may be expressed in temporal prepositions, such as, "while", "before", "after", etc.' (Qian 454, col. 11, lines 14-18; emphasis added)

Spatial descriptors similarly represent location and size information related to objects. (Qian 454, col. 11, lines 11-14) The objects are named by object descriptors:

'The object descriptors indicate the existence of certain objects in the video frame; for example, "animal", "tree", "sky/cloud", "grass", "rock", etc.' (Qian 454, col. 11, lines 8-11)

The descriptors of Qian 454 summarize the content of a shot. The metrics of Claim 1 relate to segments not objects:

"metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair".

The disclosed temporal descriptors: "while", "before", and "after" are not metrics indicating a temporal separation or accumulated duration.

Qian 454 also discloses "Hunt Information" in Figure 9, but teaches that this binary (true/false) information is only used to determine whether a valid hunt has been detected. (Qian 454, col. 11, lines 42-50) Ratakonda does not add to the teachings of Qian 454 in regard to the metrics.

The cited references do not teach or suggest generating an inter-segment temporal relationship feature of each pair of segments that includes the temporal separation and accumulated duration metrics.

Claim 1 also requires that the inter-segment temporal relationship feature includes metrics of temporal separation between the segments of the respective said pair and accumulated duration of the segments of the respective said pair. This inclusion is not taught by the cited references, nor has any motivation been proposed for one of skill in the art to make a combination of references having such a feature. The rejection indicates that an inter-segment temporal relationship feature (hereafter "ISTR feature") is not taught by Qian 454, but is taught by Ratakonda Figures 1, 5 and corresponding text. It is not apparent what the rejection proposes is an inter-segment temporal relationship feature in Ratakonda. Figure 1 shows three nested ovals, labeled "FINEST SUMMARY", "COARSE SUMMARY", and "MOST COMPACT SUMMARY". Hierarchical, multi-level summarization is discussed. (Ratakonda, col. 2, lines 61-62; col. 3, line 31 to col. 4, line 34) The word "time" is used once in this section, but that usage is meaningless to the issues here:

"... may provide a detailed fine-level summary with sufficiently large number of frames, so that important content information is not lost, but at the same time provide less detailed summaries ..." (Ratakonda, col. 3, lines 32-35; emphasis added)

Figure 5 of Ratakonda shows a hierarchical summary of keyframe levels. (see Ratakonda, col. 3, lines 1-2; col. 5, lines 44-63; col. 13, lines 27-35) "Parent" and "child" keyframes are discussed, but these terms do not denote a temporal relationship, since all of the keyframes are from video segments and the parent keyframes are not selected based on sequence order. (Ratakonda, col. 2, lines 12-26) Ratakonda does determine keyframes using an "action measure" and a "cumulative action measure" that relate to histograms of image content. (Ratakonda, col. 6, lines 25-44) If these action measures are proposed to be the inter-segment temporal relationship feature, then how would the descriptors of Qian 454 be included in the action measures of Ratakonda? The shot summaries of Qian 454 have descriptors that are textual and can be read by humans. (Qian 454, col. 10, line 63 to col. 11, line 6) The action measures of Ratakonda are defined by mathematical equations. (Ratakonda, col. 6, lines 25-44) What would lead one of skill in the art to have a reasonable expectation of the combination being a success?

The cited references do not teach or suggest merging video segments with a merging criterion that applies a probabilistic analysis of inter-segment features of pairs of segments.

Claim 1 requires:

"d) merging video segments with a merging criterion that applies a probabilistic analysis to the features of the feature set, thereby generating a merging sequence representing the video structure".

The rejection proposes that preparation of textual summaries in Qian 454 meets this language:

"merging the video segments by applying a probabilistic analysis to the extracted set to represent the video structure ("each shot is summarized 16 ...events 22 are inferred from the shot summaries by a domain specific event inference model". Column 3, lines 6-8)."

Qian 454 contradicts the rejection. Qian 454, in discussing an animal hunt example, describes various items of "hunt information" and states:

'This "hunt information" is designated true (1) or false (0) and is used in the event inference module to determine whether a valid hunt has been detected.' (Qian 454, col. 11, lines 48-50; also generally see col. 11, line 7 to col. 12, line 9)

This is not a probabilistic analysis. Ratakonda does not add to the teachings of Qian 454 in relation to this feature.

Ratakonda does not teach merging segments, but rather clustering keyframes. (Ratakonda, col. 9, lines 40-64) As a part of the process, Ratakonda also prunes keyframes of some shots. (Ratakonda, col. 8, lines 31-34) As noted above, a segment can have multiple keyframes. Ratakonda indicates:

'The number of keyframes allocated to a particular shot 's', block 42, is proportionate to the relative amount of cumulative action measure within that shot.' (Ratakonda, col. 6, lines 42-44)

This is unlike Claim 1 and contrary to the rejection.

The rejection fails to state a prima facie rejection, since the rejection fails to address rebuttal evidence showing motivation to not combine the references.

The rejection argues as motivation for one of ordinary skill in the art to combine Qian 454 and Ratakonda:

"It would have been obvious to one of ordinary skill in the art, having the teachings of Qian et al. and Ratakonda before him at the time the invention was made, to modify the segment generation and merging techniques taught by Qian et al. to include the processing of each pair of segments of Ratakonda, in order to obtain not only frames, but also inter-segment similarity processing. One would have been motivated to make such a combination because layered hierarchical structure would have been obtained, as taught by Ratakonda." (*The rejection confuses "segments" and "frames" in arguing motivation for the rejection. Contrary to this statement, frames are not obtained from segment generating and merging techniques. A shot or segment is comprised of a plurality of individual frames. (See above quote of Qian 454, col. 3, lines 37-40) For the sake of avoiding delay in prosecution, it is assumed that the words "frames" in the preceding quote from the rejection should have been "segments".*)

In the last Amendment (filed on or about Mar. 6, 2006), Applicants presented evidence in the cited references, denying the proposed motivation and instead providing motivation for one of skill in the art to not combine those references. The Office Action responded:

"In response to the arguments regarding motivation to combine Qian and Ratakonda the examiner disagrees. Ratakonda clearly teaches layers of a parent/child hierarchy structure and describes the advantages of such structure (Col. 1, line 64 et seq.)."

This citation does not answer the evidence in the cited references that teaches against the cited combination of references. The cited portion of Ratakonda describes the use of multiple keyframes to summarize a shot:

'A major limitation of the above schemes is that they treat all shots equally. In most situations it might not be sufficient to represent the entire shot by just one frame. This leads to the idea of allocating a few keyframes per each shot depending upon the amount of "interesting action" in the shot. The current state of the art video browsing systems thus split a video sequence into its component shots and represent each shot by a few representative keyframes, where the representation is referred to as "the summary". (Ratakonda, col. 1, line 64 to col. 2, line 1)

How would this be an advantage to a combination of Qian 454 with Ratakonda, unless the combination used keyframes? Such a combination would not teach or suggest the extracting and merging steps of Claim 1.

In addition, a combination of Qian 454 with Ratakonda is taught against by those references. Ratakonda teaches a "summary" in the form of keyframes of the video. (Ratakonda, col. 2, lines 13-17) In the hierarchy of Ratakonda, each level has less of the keyframes:

"Video summarization" refers to determining the most salient frames of a given video sequence that may be used as a representative of the video. A method of hierarchical summarization is disclosed for constructing a hierarchical summary with multiple levels, where levels vary in terms of detail (i.e., number of frames). The coarsest, or most compact, level provides the most salient frames and contains the least number of frames." (Ratakonda, col. 2, lines 27-33)

This "summary" of Ratakonda is not compatible with the "summary" of Qian 454, which is textual and lacks details:

"Each shot detected or forced at the first level 4 of the video content analysis technique is summarized 16 at the second level 12 of the technique. The shot summaries provide a means of encapsulating the details of the feature and motion analysis performed at the first 4 and second 12 levels of the technique so that an event inference module in the third level 18 of the technique may be developed independent of the details in the first two levels. The shot summaries also abstract the lower level analysis results so that they can be read and interpreted more easily by humans. This facilitates video indexing, retrieval, and browsing in video databases and the development of algorithms to perform these activities." (Qian 454, col. 10, line 63 to col. 11, line 6; emphasis added)

Ratakonda teaches a hierarchy, in which every level contains detail in the form of one or more keyframes. Qian 454 teaches to the contrary that detail is to be encapsulated in text, which can be read and interpreted more easily by humans.

Claims 3-5 and 23-26

Claims 3-5 and 23-26 are allowable as depending from Claim 1 and as follows.

As to Claim 3, the rejection stated:

"As in Claim 3, Qian et al. teaches obtaining unstructured video frames, generating segments from the shot boundaries based on the color dissimilarity between consecutive frames, extracting a set by processing pairs of segments for their visual dissimilarity and temporal relationship by generating color histograms from the consecutive frames and from the histograms, generating a difference signal, thresholding of this signal based on a mean dissimilarity over several frames to produce a signal representative of the existence of a shot boundary (See Claim 23 rejection supra) and merging the video segments by applying a probabilistic analysis to the extracted set to represent the video structure (See Claim 1 rejection supra) and the difference signal to be based on a mean dissimilarity over several frames centered on one frame."

Claim 3 requires a difference signal is based on a mean dissimilarity determined over a plurality of frames centered on one of the consecutive frames. The rejection proposes that Qian 454 teaches:

"thresholding of this signal based on a mean dissimilarity over several frames" (emphasis added)

Applicants respectfully traverse. Claim 3 requires a dissimilarity determined over a plurality of frames centered on one of the consecutive frames. Qian 454 teaches use of a difference in histograms between a pair of frames. Qian 454 states:

"A color histogram technique may be used to detect the boundaries of the shots. The difference between the histograms of two frames indicates a difference in the content of those frames. When the difference between the histograms for successive frames exceeds a predefined threshold, the content of the two frames is assumed to be sufficiently different that the frames are from different video shots. Other known techniques could be used to detect the shot boundaries." (Qian 454, col. 3, lines 40-50; emphasis added)

The combination of the cited references adds nothing to Qian 454, since Ratakonda is similar to Qian 454:

"Image color histograms, i.e., color distributions, constitute representative feature vectors of the video frames and are used in shot boundary detection 38 and keyframe selection. Shot boundary detection 38 is performed using a threshold method, where differences between histograms of successive frames are compared." (Ratakonda, col. 4, lines 48-54; emphasis added)

The emphasized language in the above quotes contradicts the rejection.

The office action stated in relation to Claim 5:

"As in Claim 5, Qian et al. teaches computing a mean color histogram for each segment and a visual dissimilarity feature metric from the difference between mean color histograms for pairs of segments (Column 3, lines 42-50 and Figure 5)."

"In response to the arguments regarding claim 5 and 6, Ratakonda teaches processing each pair of segments for dissimilarity in the same way Qian does for frames as seen supra."

Claim 5 requires computing a mean color histogram for each segment and computing a visual dissimilarity feature metric from the difference between mean color histograms for pairs of segments. Since Claim 5 depends from Claim 1, it also requires generating video segments from the unstructured video by detecting shot boundaries based on color dissimilarity between consecutive frames. There are, thus, two different types of dissimilarity features required by Claim 5: a first type is between consecutive frames and a second type is between pairs of segments. There is only one such feature in Qian 454. This is apparent from the office action, which cites the same portion of Qian 454 for both Claim 1 (Column 3, lines 42-43) and Claim 5 (Column 3, lines 42-50 and Figure 5). (Figure 5 of Qian 454 relates to a "sample mean" that is unrelated to the subject matter of Claim 5. See Qian 454, col. 6, lines 16-18)

Applicants respectfully traverse the statement in the rejection that:

"In response to the arguments regarding claim 5 and 6, Ratakonda teaches processing each pair of segments for dissimilarity in the same way Qian does for frames as seen supra."

Ratakonda and Qian 454 both teach a color dissimilarity feature for a pair of frames not a pair of segments. Ratakonda states:

"Image color histograms, i.e., color distributions, constitute representative feature vectors of the video frames and are used in shot boundary detection 38 and keyframe selection. Shot boundary detection 38 is performed using a threshold method, where differences between histograms of successive frames are compared." (Ratakonda, col. 4, lines 48-54)

The cited combination of references only teaches determining a color dissimilarity feature between frames.

Claim 24 has language like Claim 28 and is also allowable on the same grounds as that claim (discussed below). Claims 25-26 have language taken from Claims 7-8 (discussed below) and are also allowable on the same grounds as those claims.

Claim 7

The rejection stated in relation to Claim 7.

"As in Claim 7, Qian et al. teaches the method of claim 1 as seen supra, and generating parametric mixture models (summaries created

by shot summarization 16, Figure 1) to represent class-conditional densities of inter-segment features (based on temporal information and color analysis, See Claim 1 rejection supra) of the feature set, parametric mixture models being statistical models (Col. 3, lines 34-35, and Col. 4, lines 30 et seq.) and applying the merging criterion to the parametric mixture models (event inference 20/detected events 22, Figure 1)."

Claim 7 requires "extracting a feature set by processing pairs of said segments". As discussed in relation to Claim 1, the cited references not only do not teach or suggest, but rather teach against extracting inter-segment features by processing pairs of segments. In Qian 454, shots are compared in the form of summaries. Each of the individual shots, in Qian 454, are summarized with descriptors, such as "animal" and "tree", and the descriptors of different shots are compared, but not in pairs. (Qian 454, col. 10, line 61 to col. 12, line 9) Qian 454 teaches against comparisons between shots based upon "details" and teaches against presentation of image content to users. Qian 454 instead presents summaries to be read and interpreted. (Qian 454, col. 10, line 63 to col. 11, line 6) As discussed above, Figures 1 and 5 and related text of Ratakonda does not teach or suggest "extracting a feature set by processing pairs of said segments". Figure 5 of Ratakonda does not represent a comparison of shots, with each shot having a representative keyframe, since Ratakonda teaches against use of one keyframe per shot:

"In most situations it might not be sufficient to represent the entire shot by just one frame." (Ratakonda, col. 1, lines 65-66)

'Given total number of keyframes (user specified) 40, each shot is assigned a number of keyframes 42 depending upon the "action" within the shot, according to well known techniques.' (Ratakonda, col. 4, lines 54-57; emphasis added; also see Ratakonda Figure 2)

The cited references teach nor more together than they do individually.

The rejection also proposes that "said parametric mixture models being statistical models" is taught at Qian 454, col. 3, lines 34-35, and col. 4, lines 30 et seq. The first citation is presumed to be a phrase in Qian 454, col. 3, at lines 33-35, which states:

"but the technique may be easily extended to other domains by adding modules containing rules specific to the additional domain."

On its face, this language appears to present an inappropriate rejection in that it conveys the meaning that the claimed invention would have been well within the ordinary skill of the art at the time the claimed invention was made, because the references relied upon teach that all aspects of the claimed invention were individually known in the art. (Such art is not sufficient to establish a prima facie case of obviousness without some objective reason to combine the teachings of the references. MPEP 2143.01) The second citation from Qian 454 (col. 4, lines 30 et seq.) is indefinite in length, but is believed to refer to Qian 454, col. 4, lines 32-60, which discuss use of a five level pyramid technique to estimate global motion. (See Qian 454, col. 4, lines 18-34)

The combination of the two citations, Qian 454, col. 3, lines 34-35, and col. 4, lines 30 et seq., is understood to represent the position that a "domain" of Qian 454, col. 3, lines 34-35 could be the five level pyramid technique used to estimate global motion at Qian 454, col. 4, lines 30 et seq. There is no support for this argument in the cited references.

Qian 454 uses the term "domain" in a specific sense that is contrary to the rejection. Qian 454 states:

"The third category of techniques for analyzing video content applies rules relating the content to features of a specific video domain or content subject area. For example, methods have been proposed to detect events in football games, soccer games, baseball games and basketball games. The events detected by these methods are likely to be semantically relevant to users, but these methods are heavily dependent on the specific artifacts related to the particular domain, such as editing patterns in broadcast programs. This makes it difficult to extend these methods to more general analysis of video from a broad variety of domains.

"What is desired, therefore, is a method of video content analysis which is adaptable to reliably detect semantically significant events in video from a wide range of content domains." (Qian 454, col. 1, lines 48-62; emphasis added)

"As a result, the technique of the present invention detects event which are meaningful to a video user and the technique may be extended to a broad spectrum of video domains by incorporating shot summarization and event

inference modules that are, relatively, specific to the domain or subject area of the video which operate on data generated by visual analysis processes which are not domain specific." (Qian 454, col. 2, lines 7-18; emphasis added)

"Event inference 20 is based on domain or subject matter specific knowledge developed from observation of video and shot summaries generated at the intermediate level 12 of the technique. For example, an animal hunt usually comprises an extended period during which the animal is moving fast, followed by the slowing or stopping of the animal." (Qian 454, col. 11, lines 52-58; emphasis added)

"At the third level of the process, event inference modules provide the domain specific structure necessary for reliable event detection, but the technique may be easily extended to other domains by adding modules containing rules specific to the additional domain." (Qian 454, col. 3, lines 31-35; emphasis added)

(Qian 454 at col. 9, lines 15-22, also mentions the spatial domain relative to the spatial-frequency decompositions of Gabor filters. The rejection's proposed use of "domain" does not meet this alternative definition. In the following discussion, the "spatial domain" is not considered.)

As the emphasized language in the above quotes indicates, the term "domain" in Qian 454 is the subject area of the content of a video. An example of a domain, discussed extensively in Qian 454, is an animal hunt. The five level pyramid technique used to estimate global motion in Qian 454 is not an example of a domain.

Qian 454 also specifically teaches that the five level pyramid technique is not specific to a particular domain. Qian 454 states:

"At the lowest level of the technique, visual analysis processes which are not specific to an application or video domain provide basic information about the content of the video." (Qian 454, col. 3, lines 17-20; emphasis added; see also col. 2, lines 11-18)

"At the first level 4 of the technique, the global motion of the video content is estimated 8 for each pair of frames in a shot." (Qian 454, col. 3, lines 59-61; emphasis added)

"In the instant technique, global motion is estimated with a five level pyramid technique". (Qian 454, col. 4, lines 18-19; emphasis added)

The emphasized language shows, the position taken in the rejection is unsupported by the cited references, since the five level pyramid technique for estimating global motion is not domain specific.

The rejection proposes that creating summaries in Qian 454 by shot summarization based on temporal information and color analysis, corresponds to generating parametric mixture models to represent class-conditional densities of inter-segment features of a feature set extracted by processing pairs of segments. The rejection also proposes that applying the merging criterion to the parametric mixture models is taught in Qian 454 by "(event inference 20/detected events 22, Figure 1)". The cited combination of references do not disclose these features.

The cited references do not teach or suggest generating parametric mixture models to represent class-conditional densities of inter-segment features of the feature set. Claim 7 requires generating "parametric mixture models" that are defined by the specification and usage in the art as types of statistical models. (See application page 4, lines 25-30; page 13, lines 14-29; also see U.S. Patent No. 5,710,833.) As discussed above, the cited references do not teach or suggest this. The rejection's "summaries created by shot summarization" are not statistical models. Qian 454 teaches summaries, in which shot descriptors are described as indicating as to a particular shot: "the existence of certain objects", "location and size information related to objects and the spatial relations between objects", and "motion information related to objects and the temporal relations between them". (Qian 454, col. 11, lines 9, 11-13, and 15-16; see also the above discussion of summarization.)

In Claim 7, the parametric mixture models are generated to represent class-conditional densities of inter-segment features. The cited references, as noted above, fail to teach or suggest the processing of pairs of segments to provide the features of the feature set. Shot descriptors for each segment are taught by Qian 454. (See Qian 454, col. 10, lines 61-62: "Each shot ... is summarized"). Ratakonda teaches a "summary" in the form of a hierarchy of keyframes, which is not limited to one keyframe per shot. (Ratakonda, Figures 1

and 5, col. 3, lines 30-45; col.5, lines 44-63; and col. 13, lines 22-31; col. 1, lines 65-66; col. 4, lines 54-57)

The rejection is not supported by the cited references and must be withdrawn.

Claims 6 and 8 are allowable as depending from Claim 7 and as follows.

Claim 6

The rejection stated in relation to Claim 6:

'As in Claim 6, Qian et al. teaches processing pairs of segments for a temporal separation between pairs of segments and for an accumulated temporal duration between pairs of segments ("each shot is summarized 16 ... events 22 are inferred from the shot summaries by a domain specific event inference model". Column 3, lines 6-8).'

'In response to the arguments regarding claim 5 and 6, Ratakonda teaches processing each pair of segments for dissimilarity in the same way Qian does for frames as seen supra.'

Claim 6 is allowable on the same grounds as discussed above in relation to similar language in Claim 1. Claim 6 requires that the processing of pairs of segments includes processing for a temporal separation between pairs of segments and for an accumulated temporal duration of pairs of segments. The language relied upon in the rejection (Qian 454, col. 3, lines 6-8) relates to inferences based on textual shot summaries. Qian et al. states:

"The shot summaries also abstract the lower level analysis results so that they can be read and interpreted more easily by humans." (Qian 454, col. 11, lines 1-3)

Qian 454 teaches summarization that encapsulates the details of the feature and motion analysis of each shot using descriptors. (Qian 454, col. 10, line 63 to col. 11, line 8) The domain specific event inference model uses the descriptors. (Qian 454, col. 11, lines 51-55) Events are inferred by matching the occurrence of objects and their spatial and temporal relationships detected in each of the shots. (Qian 454, col. 12, lines 6-7; generally see col. 11, line 58 to col. 12, line 9) Examples of shot descriptors are provided:

'In general, shot descriptors used in the shot summary include object, spatial, and temporal descriptors. The object descriptors

indicate the existence of certain objects in the video frame; for example, "animal", "tree", "sky/cloud", "grass", "rock", etc. The spatial descriptors represent location and size information related to objects and the spatial relations between objects in terms of spatial prepositions, such as "inside", "next to", "on top of", etc. Temporal descriptors represent motion information related to objects and the temporal relations between them. These may be expressed in temporal prepositions, such as, "while", "before", "after," etc.' (Qian 454, col. 11, lines 7-18)

Qian 454 does not teach descriptors of temporal separation between pairs of segments and/or for accumulated temporal duration between pairs of segments. In Qian 454, temporal descriptors represent motion information related to objects in a segment and the temporal relations between those objects in that segment. This is unlike Claim 6, which requires processing pairs of segments for a temporal separation between pairs of segments and for an accumulated temporal duration of pairs of segments.

Ratakonda does not disclose the features of Claim 6. Ratakonda instead teaches the use of histogram clustering to determine keyframes. (See Ratakonda, col. 9, lines 30-53, quoted above)

Claim 8

The Office Action stated in relation to Claim 8:

'As in Claim 8, it is notoriously well known that queues are used to implement hierarchical displays. The examiner takes official notice of this teaching. It would be obvious to one of ordinary skill in the art to combine the use of the organizing video segments into hierarchies with a queue implementation.'

The Office Action presents a different rejection of Claim 15, which is very similar to Claim 8:

"As in Claim 15, US Patent 6721454 and Ratakonda teach performing the merging in a hierarchical queue by initializing the queue by introducing each feature in the queue with a priority of the probability of merging each corresponding pair of segments, depleting the queue by merging the segments if the criterion is met, and updating the queue based on the updated model (See Claim 8 rejection supra)."

The Office Action also stated:

'In response to the arguments regarding claim 8 and 15, Qian teaches the process of "inserting" merges frames together, constituting a pair of segments that define the event and updating the model of the merged segment. Ratakonda further illustrates step d as seen supra. Furthermore, claim 15 is interpreted with respect to the official notice of Claim 8. Claim 15 merging and depleting sequence can further be illustrated by Qian figures 1 and 7 with corresponding text.'

In the previous amendment filed on or about March 6, 2006, in relation to Claim 8, Applicants presented a demand for clarification of the official notice stating:

"Clarification of the rejections of Claims 8 and 15 is requested, particularly as to the metes and bounds of the official notice taken and of the relied upon teachings of Qian 454 and Ratakonda."

The Office Action presents a response that is inadequate. The metes and bounds of the official notice taken are not addressed. Applicants have insufficient information to prepare a response to the rejection. The rejection will not stand.

The Office Action does present additional argument as to the references, but in so doing again confuses the meanings of "frames" and "segments", stating:

"Qian teaches the process of "inserting" merges frames together, constituting a pair of segments that define the event and updating the model of the merged segment." (emphasis added)

As discussed above, frames are not segments. Thus, merging frames together cannot constitute "a pair of segments that define the event and updating the model of the merged segment."

The Office Action indicates "Qian teaches the process of "inserting" merges frames together". It is unclear what is meant by this. Qian 454 mentions "inserting", but this term has the same meaning as "forcing", that is, adding another shot boundary into a sequence of frames. Qian 454 states:

"In addition to the shot boundaries detected in the video sequence, shot boundaries may be forced or inserted into the sequence whenever the global motion of the content changes. As a result, the global motion is relatively homogeneous between the boundaries of a shot. In addition, shot boundaries may be forced after a specific number of frames (e.g., every

200 frames) to reduce the likelihood of missing important events within extended shots." (Qian 454, col. 3, lines 51-58; emphasis added)
Inserting a shot boundary divides a video sequence. This is the opposite of merging.

The Office Action also states:

"Ratakonda further illustrates step d as seen supra."

This is understood to mean the same thing as the discussion of step d in the rejection of Claim 1:

"Ratakonda teaches a video event detection and segmentation merging method similar to that of Qian et al."

As discussed above, Ratakonda does not teach merging segments, but rather producing a hierarchy of keyframes by clustering keyframes. (Ratakonda, col. 9, lines 40-64) As a part of the process, Ratakonda also prunes keyframes of some shots. (Ratakonda, col. 8, lines 31-34) A segment can have multiple keyframes. (Ratakonda, col. 6, lines 42-44)

The rejection of Claim 8 is otherwise limited to the words of the official notice. The rejection states that it is notoriously well known that queues are used to implement hierarchical displays. This statement addresses only one phrase of Claim 8: "performed in a hierarchical queue" and does not teach or suggest the steps of:

initializing the queue by introducing each feature into the queue with a priority equal to the probability of merging each corresponding pair of segments;

depleting the queue by merging the segments if the merging criterion is met; and

updating the model of the merged segment and then updating the queue based upon the updated model.

The rejection will not stand and must be withdrawn.

Claim 10

Claim 10 is supported and allowable on the grounds discussed in relation to Claim 7.

Claims 27-29

Claim 27 requires "said inter-segment temporal relationship feature including metrics of temporal separation between the segments of the respective

said pair and accumulated duration of the segments of the respective said pair" and is allowable on grounds discussed above in relation to Claim 1.

Claim 28 is allowable as depending from Claim 27 and requiring metrics related to numbers of frames. This is not possible if the meanings of "segment" and "frame" are equated.

Claim 29 requires "said extracting of said inter-segment temporal relationship feature of each said pair of segments including determining a number of frames separating the respective said pair of segments and determining an accumulated number of frames in said segments of the respective said pair of segments" and is allowable on the same grounds as Claims 27-28.

Claims 9 and 11-22 stand rejected under 35 U.S.C. 103(a) as being unpatentable over Qian et al., US Patent 6721454 (hereafter (Qian 454) and Ratakonda, US Patent 5956026 and further in view of Qian et al., US Patent 6616529 (hereafter Qian 529). The rejection stated:

'As in Claims 9, 11, 17-18 and 20, US Patent 6721454 and Ratakonda teach a method and computer storage medium with instructions for obtaining unstructured video frames, generating segments from the shot boundaries based on the color dissimilarity between consecutive frames, extracting a set by processing pairs of segments for their color dissimilarity and temporal relationship of each pair of segments, merging adjacent video segments by applying a probabilistic analysis to the extracted set to represent the video structure, and generating a parametric mixture model of the inter-segment features (See Claim 1 rejection supra), the parametric mixture models being a statistical model (Col. 3, lines 34-35, and Col. 4, lines 30 et seq.). While US Patent 6721454 and Ratakonda teach the segmentation due to color dissimilarity, extraction due to visual dissimilarity and temporal relationships, merging with probabilistic analysis and generation of a parametric mixture model, they fail to show the probabilistic analysis to be a Bayesian analysis applied to the parametric mixture model, and representing the merging sequence in a hierarchical tree structure as recited in the claims. US Patent 6616529 teaches a video segmentation method similar to that of US Patent 6721454 and Ratakonda. In addition, US Patent 6616529 further teaches the

probabilistic analysis to be a Bayesian analysis applied to the parametric mixture model (Figure 3 and corresponding text in Columns 4-5), and representing the merging sequence in a hierarchical tree structure (Figures 2a-2g and corresponding text). It would have been obvious to one of ordinary skill in the art, having the teachings of US Patent 6721454 and Ratakonda and US Patent 6616529 before him at the time the invention was made, to modify the segmentation with color dissimilarity and temporal relationships with a parametric mixture model taught by US Patent 6721454 and Ratakonda to include the construction of hierarchy according to probabilistic merging with Bayesian analysis of US Patent 6616529, in order to obtain a hierarchical representation of the frames grouped by color dissimilarity and temporal relationships according to Bayesian probability methods of analysis. One would have been motivated to make such a combination because a visual representation of the segmented video would have been obtained, as taught by US Patent 6616529 (Column 2, lines 24-55).'

Claim 9 is allowable as depending from Claim 1.

Claim 11 is allowable as discussed above in relation to Claim 7.

Claims 12-16 are allowable as depending from Claim 11. Claims 12-13 are also allowable on the same basis as Claims 5-6, respectively. Claim 15 is allowable on the same basis as Claim 8.

Claims 17-18 are allowable as discussed above in relation to Claim

7.

Claim 19 is allowable as depending from Claim 18.


Claim 20 is allowable as discussed above in relation to Claim 7.

Claim 21 is allowable on the same basis as Claim 1.

Claim 22 is allowable as depending from Claim 21.

In view of the foregoing and previous distinctions made in earlier responses all of the claims are patentably distinct from the prior art. None of the references, taken singly or in combination, disclose the claimed invention. Accordingly, this application is believed to be in condition for allowance, the notice of which is respectfully requested

Respectfully submitted,

A handwritten signature in black ink, appearing to read "Robert Luke Walker", written over a horizontal line.

Attorney for Applicant(s)
Registration No. 30,700

Robert Luke Walker/ld
Rochester, NY 14650
Telephone: (585) 588-2739
Facsimile: (585) 477-1148